

General

DMAC ID: 6.2

Suggested Name: Standards package for the representation and transport of gridded data: netCDF+CF+OPeNDAP+aggregation

Submission Date: 10/30/2007 07:17 PM

Type of Standard, guidance or best practice (select one).
New

Data Management Function(s)
Data Transport

Summary

Abstract:

The combination of the netCDF API (v 3.6), CF (v1.0) and OPeNDAP (DAP2) with time aggregations provides a foundation for the representation and transport of gridded model output, bin-averaged climatological collections, and gridded satellite fields. The CF specification includes controlled vocabularies and detailed encodings for a broad range of structural semantics for grids. NetCDF provides a persistence format for archive and a high performance access mechanism suitable for local file-based applications. OPeNDAP with time-aggregation (which unifies collections of netCDF files along their time axis into logical data sets of greater scope than the individual files) provides a mechanism for efficient dissemination of the data via Web services.

This package of four components is widely used in those quarters of the marine science community that work with multi-variate, multi-dimensional gridded fields, as well as strong interoperability partnerships with atmospheric and climate sciences. It has a proven track record over many years, many institutions, and many terabytes of data. It is supported by a vigorous international community of developers, data suppliers and users. This is a standard of sufficient breadth, specificity and robustness to be immediately usable by the IOOS community.

Note: This submission has a further goal, beyond the certification of a useful standard (or rather standards package) within the IOOS/DMAC data management community. Namely, this submission is also intended to explore the relative merits of applying the DMAC standards process to application-ready collections of standards (such as this) versus single component standards (such as DAP 2, alone -- a component of this standard which is also being proposed independently.)

Purpose/Scope:

The combination of the netCDF API (v 3.6), CF (v1.0) and OPeNDAP (DAP2) with time aggregations provides a foundation for the representation and transport of gridded model output, bin-averaged climatological collections, and gridded satellite fields. Such data types in their most general description are multi-variate (commonly up to 50 variables); multi-dimensional (lat, long, depth, and time) with different variables possessing different dimensionality – e.g. ocean surface fields mixed with height/depth varying fields; and with a variety of spatial coordinate structures (regular rectilinear, irregular rectilinear, curvilinear, fixed-Z, sigma-Z, isopycnal-Z, hybrid-Z). The CF specification includes controlled vocabularies for parameter names, and classification of sub-grid cell analyses. OPeNDAP provides remote access to these data over standard HTTP through applications that utilize the standard netCDF 3.6 API. Server-side subsetting capabilities, that are made available through the netCDF API and OPeNDAP, offer multiple dramatic performance advantages over file transfers, particularly for large files. Time-aggregation (presenting a time-sequence of files as a single virtual file) greatly simplifies the process of data navigation for users and simplifies the process of extracting useful data structures, such as time series that may span multiple files. [Note that netCDF files created with the netCDF 3.6 API that follow CF 1.0 conventions may provide a self-describing, persistence format suitable for archive and for high performance local file-based applications with random access (subsetting), although standardizing archive formats is not a specific intent of this standards profile.]

Limits to the scope of this standards package:

The package of standards described here is the product of a community of software architects, software developers, data providers and data users. This community has already advanced the functionality of the standards named herein well beyond what is proposed in this package for prompt standardization. The endusers of this package fully expect that there will be a continual roll-out and standardization of new capabilities from this package. However, at this time we are putting forward only a stable core of capabilities. Please see the section entitled "Technical Description" for a list of development areas that recognized as important, but are not currently ready for standardization in IOOS.

Technical Description:

Brief technical summaries of the components:

NetCDF

NetCDF (network Common Data Form) is a set of interfaces for array-oriented data access and a freely-distributed collection of data access libraries for C, Fortran, C++, Java, and other languages. The netCDF libraries support a machine-independent format for representing scientific data. Together, the interfaces, libraries, and format support the creation, access, and sharing of scientific data.

NetCDF data is:

- Self-Describing. A netCDF file includes information about the data it contains.
- Portable. A netCDF file can be accessed by computers with different ways of storing integers, characters, and floating-point numbers.
- Direct-access. A small subset of a large dataset may be accessed efficiently, without first reading through all the preceding data.
- Appendable. Data may be appended to a properly structured netCDF file without copying the dataset or redefining its structure.
- Sharable. One writer and multiple readers may simultaneously access the same netCDF file.
- Archivable. Access to all earlier forms of netCDF data will be supported by current and future versions of the software.

CF

The CF conventions for climate and forecast metadata are designed to promote the processing and sharing of files created with the NetCDF API

The conventions define metadata that provide a definitive description of what the data in each variable represents, and of the spatial and temporal properties of the data. This enables users of data from different sources to decide which quantities are comparable, and facilitates building applications with powerful extraction, regridding, and display capabilities.

The CF conventions generalize and extend the COARDS conventions.

OPeNDAP

OPeNDAP: Open-source Project for a Network Data Access Protocol

OPeNDAP provides software which makes local data accessible to remote locations regardless of local storage format. OPeNDAP also provides tools for transforming existing applications into OPeNDAP clients (i.e., enabling them to remotely access OPeNDAP served data). The foundation of OPeNDAP is the DAP – the Data Access Protocol. Remote access to netCDF (and other format) data through the netCDF API carried over the DAP is a default capability of the Java netCDF libraries as distributed by Unidata and is available for the C libraries from OPeNDAP, Inc. (Note: Funded work is underway that will blend the C library support for OPeNDAP into the default C libraries supplied by Unidata, as well.) A number of other language bindings for netCDF are derived from the C language support – FORTRAN, perl, Python, etc.

Time-aggregation

Time-aggregation refers to the practice of presenting a time-sequence of files as a single virtual netCDF file through the OPeNDAP interface. Strictly-speaking aggregation is a "best practice" rather than a standard; the virtual netCDF-CF datasets assembled through the process of aggregation should be 100% adherent to the specifications of the netCDF API and the CF conventions – no new standards compliance issues are raised.

Several tools exist to create virtual time-aggregations from sequences of files: the THREDDS Data Server (TDS) from Unidata, the GrADS/DODS Server (GDS) from COLA and the HYRAX server from OPeNDAP, Inc. Access to virtual aggregations and

server-side subsetting capabilities via OPeNDAP greatly reduces the system resources required to serve and utilize tera-scale model outputs. Through these tools clients are able to request and receive surgical subsets, from much larger data collections. Such an approach commonly reduces the download requirements by factors of 10^4 or more.

Server-side aggregation and OPeNDAP access also make it possible to present virtual CF file access from collections of files that are not natively encoded in netCDF. Such files include the WMO GRIB-formatted files that are routine in WMO operations and (with limitations) the HDF files that are commonly found in satellite data management. Thus the standards proposed herein provide a pathway to interoperable data exchange between neighboring communities. Note: see the section on Citation/provenance under "Limits to the scope of this standards package" for important considerations regarding the use of aggregation.

**** Current limitations to the scope of this standards package ****

A number of capabilities exist (at varying levels of development) within the most recent versions of the standards proposed within this package, but will not be not regarded as sufficiently mature to proposed for standardization at this time. These capabilities include :

Tiled datasets (also called "gridspec" files) – Additional metadata and coordinate data encoded in CF to describe certain complex coordinate reference systems such as the Cube Sphere.

Web Coverage Service (WCS) interoperability – Through the GALEON/OGC process much of the netCDF-CF 1.0 standards combination has been endorsed by the OGC and made available through the WCS standard. There is a high level of overlap and interoperability between the package proposed herein and the work of GALEON

Citation/provenance – When on-disk "files" are presented as "virtual files" through OPeNDAP services, the ability to provide a guaranteed citation back to the original dataset is lost. This problem is aggravated by the process of aggregation, through which a single virtual file, may actually draw data from multiple physical (citable) files. This problem is not unique to OPeNDAP access. It has close parallels in Web Service access to relational database contents. But it is important to recognize that OPeNDAP and server-side aggregation require further work before they are ready to be proposed as standards in environments where strict citations to data archives are required.

Ensemble and forecast axes – Server-side aggregation techniques are applicable to assemble collections of 4-dimensional files into five and six dimensional virtual datasets, in which forecast time and model ensemble membership are presented as two additional netCDF dimensions. While these capabilities may be vital for future interoperability, they are not sufficiently mature to propose for standardization at this time.

OGC metadata compatibility -- At the time of this writing work is underway to harmonize the CF standard with metadata elements from the OGC geospatial standards specifications. These elements include representation of the geoid, named map projections, and vertical datum choices, among others.

Access control – The current implementations of OPeNDAP (and the DAP2 standard) contain only limited forms of access control (username/password). Work to extend this is underway, but is not sufficiently mature to propose in this standards package.

Common data model – Efforts are underway to unify the representation of gridded data in CF with collections of in-situ observations and unstructured meshes within a single Common Data Model (CDM). While this work is likely to play a key role in future versions of this standards package, it is not sufficiently mature to propose at this time.

Additional Information:

Statutory Requirement:
No

Relationship/Dependency/Conflicts:

Current Usage

A partial list of marine and climate data sites utilizing the proposed standards package:

- o NOAA:
- o GFDL
- o NCEP NOMADS service
- o PMEL
- o FSL
- o NCDC NOMADS service
- o NDBC (National Buoy Data Center)
- o NOS/COOPS
- o NMFS/PFEL
- o ESRL (formerly CDC)
- o NODC (under development)
- o DOE/PCMDI
- o NASA/JPL
- o NASA/GSFC
- o US Navy/ Fleet Numerical/GODAE Data Server
- o National Snow and Ice Data Center (NSIDC)
- o USGS, Woods Hole Field Center (WHFC)
- o National Center for Atmospheric Research (NCAR)
- o Center for Ocean-Land-Atmosphere Studies (COLA)
- o George Mason University (GMU) Center for Earth Observing and Space Research
- o Asia-Pacific Data-Research Center (APDRC)
- o (Australia) CSIRO Marine and Atmospheric Research (CMAR) Remote Sensing
- o (Australia) Centre for Australian Weather and Climate Research (CAWCR)
- o (France) AVISO (satellite data)
- o (France) French Research Institute for Exploitation of the Sea (IFREMER)
- o (Italy) International Centre for Theoretical Physics (ICTP)
- o (South Korea) Seoul National University (SNU)
- o (Netherlands) European Climate Assessment (ECA)

Justification

A high level of data interoperability between projects producing gridded data fields is vital to our ability to serve US societal needs. Model-generated and satellite-observed gridded data are vital to operational coastal ocean state estimations and forecasts; to understanding and forecasting climate; to weather prediction and distribution; and to many other areas of marine science endeavor. The needs and the technical demands are particularly acute for time-critical outputs, since model-generated data volumes may be extraordinarily large and the semantics of the data are subtle and complex. Service-oriented techniques that support surgical extraction of the desired data subsets are vital to the usability of the data. Semantically rich shared standards are vital to interoperability.

Although this technically-oriented summary does not include an economic analysis, there can be little doubt that an economic analysis would reveal a profound economic benefit to society from interoperable distribution of gridded data and a large potential gain if the level of interoperability can be increased.

References

| Name | URL |
|--|---|
| netCDF 3.x | http://www.unidata.ucar.edu/software/netcdf/ |
| CF 1.0 (under the authority of the WCRP) | http://cf-pcmdi.llnl.gov/ |
| OPeNDAP (DAP 2) | http://www.opendap.org/support/docs.html |
| aggregation via THREDDS | http://www.unidata.ucar.edu/projects/THREDDS/ |
| aggregation via GDS | http://www.iges.org/grads/gds/ |

Acronyms

Contact Information

| | |
|--------------------|--------------------------|
| Name: | Steve hankin |
| Organization Name: | NOAA/PMEL |
| Telephone: | (206)526-6080 |
| E-Mail Address: | steven.c.hankin@noaa.gov |

Supporting Parties/Members

| Member | Organization | E-Mail Address | Telephone Number |
|------------------|---|--|------------------|
| Dr. Rich Signell | USGS/Woods Hole & IOOS/MAST | rsignell@usgs.gov | (508) 457-2229 |
| Dr. John Wilkin | Rutgers University/Institute of Marine and Coastal Sciences & IOOS/MAST | jwilkin@rutgers.edu | (732)932-6555 |
| Ben Domenico | Unidata/UCAR (head of OGC GALEON project) | Ben@unidata.ucar.edu | (303)497-8631 |
| Dr. Frank Bub | Naval Oceanographic Office & IOOS/MAST | frank.bub@navy.mil | (228)688-4758 |
| Dr. V. Balaji | Princeton University (NOAA/GFDL) | V.Balaji@noaa.gov | (609)452-6516 |

Comments History

| | |
|----------------------|---|
| No comments provided | Status: 6.1 Action: Recommended Posted: 01/06/2009 By: Delta Pelgrim |
| Ready for poll | Status: 5.2 Action: Not Recommended Posted: 10/16/2008 By: Delta Pelgrim |
| REady for poll. | Status: 5.1 Action: Recommended Posted: 10/16/2008 By: Delta Pelgrim |

| | |
|---|---|
| No comments provided | Status: 4.1 Action: Recommended Posted: 08/19/2008 By: aluthra1 |
| No comments provided | Status: 3.6 Action: Recommended Posted: 01/22/2008 By: Delta Pelgrim |
| ready for poll | Status: 3.4 Action: Recommended Posted: 11/28/2007 By: Delta Pelgrim |
| No comments provided | Status: 3.2 Action: Recommended Posted: 11/28/2007 By: Delta Pelgrim |
| Moved to submitted by Chris Moore for Anne Ball on Medical Leave Forward to Transport expert team | Status: 2.2 Action: Recommended Posted: 11/07/2007 By: christopher moore |
| Submission complete | Status: 1.2 Action: Recommended Posted: 11/07/2007 By: Delta Pelgrim |

Public Comments History

No Comments Found

Expert Response to Public Comments History

No Comments Found